Nov 2nd 2009

Dear colleagues,

Thank you for participating of the LBL CRD Machine Learning Reading Group. We had 10 participants in the first meeting, a nice critical mass to learn and discuss about machine learning theory and applications.

Here it follows our upcoming schedule and a brief description of our first meeting:

1) Schedule:

- Mondays at 12pm to 1pm;
- next meeting: Nov 30th, 2009.

2) Introductions and interests:

- Most of the participants work with large datasets, which analysis is somewhat unfeasible without using an intelligent computer technique to prospect for subsets that follow some specific pattern. On the other hand, we must bear in mind that the time complexity of some machine learning algorithms can turn into a second problem;
- Some scientific problems require sophisticated feature extraction algorithms (e.g. PCA) before modeling prior knowledge (e.g. cybersecurity, molecule comparison) while others seem more focused on ways of dealing with the (previous extracted) features by applying clustering and/or classification algorithms (e.g. astrophysics, auto-tuning, particle simulations).

3) Homework:

- Elaborate a brief but detailed description (100-200 words) of your scientific problem(s) and machine learning techniques that are currently in use or you think may be useful; make sure you give enough details of the domain (e.g. dimensionality, feature dependence, existence of classes). Be specific about the scientific question you want to answer. For those working in more than one project, consider writing separate descriptions for each of them;
- Email descriptions of your problems to me by next *Monday, 11/09* and I'll make it available to all at;
- Read the paper <u>Statistical pattern recognition: a review</u>, AK Jain, RPW Duin, J Mao, pp.4-35, 2000 for the next meeting (http://www.mts.jhu.edu/~priebe/COURSES/FALL2003/550.730/jdm00.pdf)
- Suggestions on future papers are welcome.

Let me know if I forgot something or if you have any comments, Daniela